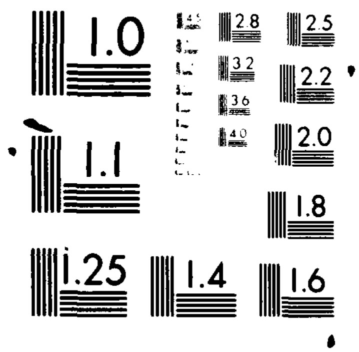
AD-A099 428   MISSOURI UNIV-COLUMBIA DEPT OF STATISTICS   F/G 12/1
IDENTIFIABILITY PROBLEMS IN THE THEORY OF COMPETING AND COMPLEM--ETC(U)
AUG 80   A P BASU   N00014-78-C-0655
UNCLASSIFIED   TR-97   NL

END
DATE
FILMED
6 81
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

AD A099428

University of Missouri-Columbia

Identifiability Problems in the Theory of Competing and Complimentary Risks— A Survey

by

Asit P. Basu
Department of Statistics

Technical Report No. 97
Department of Statistics August 1980

Mathematical Sciences

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 97	2. GOVT ACCESSION NO.: AD-A099428
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Identifiability Problems in the Theory of Competing and Complementary Risks--A Survey.	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Asit P. Basu	8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0655
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Missouri, Columbia, Missouri 65211	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia	12. REPORT DATE: Aug [redacted] 80
	13. NUMBER OF PAGES: 26 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

competing risks; complementary risks; identifiability; reliability; biometry; distribution of minimum and maximum; series and parallel system.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this expository paper the concepts of competing and complementary risks are defined and a survey of recent results in the area is presented. Identifiability of distributions, both univariate and multivariate, useful in reliability and survival analysis is considered.

AMS Classification numbers: Primary 60G05; Secondary 62N05

IDENTIFIABILITY PROBLEMS IN THE THEORY OF

COMPETING AND COMPLEMENTARY RISKS -

A SURVEY

by

Asit P. Basu*

Technical Report No. 97
August 1980

Prepared under contract N00014-78-C-0655
for the Office of Naval Research

Department of Statistics
University of Missouri
Columbia, Missouri 65211

ABSTRACT

In this expository paper the concepts of competing and complementary risks are defined and a survey of recent results in the area is presented. Identifiability of distributions, both univariate and multivariate, useful in reliability and survival analysis is considered.

Key Words: Competing risks, Complementary risks, Identifiability, Reliability, Biometry, Distribution of minimum and maximum, series and parallel system.

1. INTRODUCTION

The problem of identifiability (or of nonidentifiability) arises naturally in a number of physical situations. The problem, in general terms, can be defined as follows. Let U be an observable random variable with distribution function F_θ, and let F_θ belong to a family $F = \{F_\theta: \theta \varepsilon \Omega\}$ of distribution functions indexed by a parameter θ. Here θ could be scalar or vector valued.

We shall say θ is nonidentifiable by U if there is at least one pair (θ, θ') $\theta \neq \theta'$, θ, $\theta' \varepsilon \Omega$ such that $F_\theta(u) = F_{\theta'}(u)$ for all u. In the contrary case we shall say θ is identifiable. It may happen that θ itself is nonidentifiable but a function $\gamma(\theta)$ is identifiable in the following sense: For any θ, $\eta \varepsilon \Omega$, $F_\theta(u) = F_\eta(u)$ for all u implies $\gamma(\theta) = \gamma(\eta)$; in this case we may say θ is partially identifiable.

Puri (1979) has surveyed some examples arising in the general literature. The purpose of the present paper is to present a survey of results available in the area of competing risks and complimentary risks. We would primarily consider the cases where we have an underlying parametric model. In Sections 2 and 3 we define the problems of competing and complementary risks and discuss the case when

the underlying random variables are independently distributed. The case of dependent random variables is considered in Section 4. The problem of estimation is briefly discussed in Section 5. Finally, in Section 6, some open problems and other related areas of research are pointed out.

2. INDEPENDENT RANDOM VARIABLES WITH IDENTIFIED EXTREMUM

The problem of competing risks, in its simplest form, may be described as follows: Let X_i be a random variable with distribution function $F_i(x)$, $(i = 1,2,\ldots,p)$. We assume that X_i's are not observable, but $U = \min(X_1,\ldots,X_p)$ is. We would like to estimate F_i's given the observations on U.

This model finds interesting applications in a number of fields but particularly in problems of survival analysis and reliability theory. Thus in problems of competing risks, or in studying reliability of complex systems, an individual or a system of components may be exposed to p different causes of death (failure) where X_i is the time to death from cause C_i $(i = 1,2,\ldots,p)$. Although one would like to know about the distribution of the X_i's only observations on U's will be available. Basu and Ghosh (1980) give some of these examples. For other examples and a survey of the area see Birnbaum (1979) and David and Moeschberger (1978). An interesting ecological application is given by Anderson and Burnham (1976) who study the population ecology of the mallard where the two causes of death are "hunting" and "natural mortality."

In case the X_i's are independent and identically distributed, the problem is a trivial one since

$$\bar{F}_U(x) = \prod_{i=1}^{p} \bar{F}_i(x) = (\bar{F}_X(x))^p \tag{1}$$

where $\bar{F} = 1 - F$ is the survival function (reliability function), $F_U(x)$ is the cdf of U, and $F_X(x)$ is the common distribution function of the X_i's.

In the general case, when the X_i's are not identically distributed, the problem of identifiability can be illustrated using the following example.

Example 1.

Let X_i (i = 1,2,3,4) be independent random variables and let $X_i \sim e(\lambda_i)$. That is let the cdf of X_i be

$$F(x) = 1 - e^{-\lambda_i x}, \; x > 0, \; \lambda_i > 0.$$

If $\lambda_1 + \lambda_2 = \lambda_3 + \lambda_4$, then

$$\min(X_1,X_2) \text{ and } \min(X_3,X_4)$$

are identically distributed.

It is possible to introduce additional random variables so that the *enhanced family* with the added information becomes identifiable. In this case we call the original family of distributions "rectifiable." Let I be an integer valued random variable (I = 1,2,...,p). (U,I) is called an identified minimum if I = k when $U = \min(X_1,X_2,\ldots,X_p) = X_k$. In the absence of I, U will be called a non-identified minimum.

Basu and Ghosh (1980) introduce a dual problem, called the problem of complementary risks, where instead of observing the minimum (identified or non-identified) one observes the maximum $V = \max(X_1,X_2,\ldots,X_p)$. Basu and Ghosh (1980) introduce examples to show that this problem also occurs naturally in survival analysis and reliability theory. If the X_i's are iid then, as in the case of minimum, one can trivially obtain the cdf $F_X(x)$ of X from that of V.

We now consider the case when the X_i's are independently but not identically distributed. The joint probability distribution of (U,I) is specified by monotonic functions

$$H_k(x) = P(U \leq x,\ I = k),\ (k = 1,2,\ldots,p). \tag{2}$$

Then Berman (1963) has obtained the following theorem.

Theorem 1. The set of functions $\{H_k(x)\}$ is related to the set $\{F_j(x)\}$ by the functional equations

$$H_k(x) = \int_{-\infty}^{x} \prod_{j \neq k} (1 - F_j(t))dF_k(t),\ k = 1,2,\ldots,p. \tag{3}$$

The solution of this set of equations is

$$F_k(x) = 1 - \exp\{-\int_{-\infty}^{x} [1 - \sum_{j=1} H_j(t)]^{-1} dH_k(t)\}, \tag{4}$$

$$k = 1,2,\ldots,p.$$

Let $p_k = P(I = k)$ and $G_k(x) = \int_{-\infty}^{x} g_k(t)dt$ be the distribution function of the conditional random variable $X_k = U|I = k$. Hence

$$H_k(x) = p_k G_k(x).$$

Nádas (1970) has given an alternate expression for Theorem 1 as

$$r_k(x) = p_k g_k(x)/[\sum_{j=1}^{p} p_j(1 - G_j(x))] \tag{5}$$

and

$$F_k(x) = 1 - \exp\{-\int_{\infty}^{x} r_i(t)dt\}. \tag{6}$$

That is $F_k(x)$ are uniquely determined by the distribution of the observable pair (U,I). Where $r_k(x) = \frac{f_k(x)}{1 - F_k(x)}$ is variously called as <u>force of decrement</u> (<u>or of mortality</u>), <u>age specific death rate</u>, <u>failure rate function</u>, <u>intensity function</u>, and <u>hazard function</u>.

Since the minimum U and the maximum V are related through the relation

$$\text{Max}(X_1, X_2, \ldots, X_p) = - \text{Min}(- X_1 - X_2, \ldots, - X_p)$$

it follows that similar results also hold for the maximum.

The next natural questions are: (a) What if the minimum (maximum) is not identified? Could we obtain the distribution of X_k from that of U? And (b) What can we say about the identifiability of the F_k's if the X_k's are not independently distributed? We will discuss these next.

3. INDEPENDENT RANDOM VARIABLES WITH NONIDENTIFIED EXTREMUM

In case the extremum U is not identified, one can still uniquely determine $F_k(x)$ under certain conditions. To this end Basu and Ghosh (1980) obtain the following theorem.

<u>Theorem 2</u>. Let $\mathcal{F}$ be a family of pdf on R_1 with support (a,b) which are continuous and are positive to the left of some point A and <u>such that</u> if f and g are any two distinct members of $\mathcal{F}$ then

$$\lim_{x \to a}\{f(x)/g(x)\}$$

exists and equals either 0 or ∞. Let $X_1,\ldots,X_p$ be independent random variables with respective pdf's $f_1,f_2,\ldots,f_p$ in $\mathcal{F}$ and $Y_1,Y_2,\ldots,Y_q$ be independent random variables with respective pdf's belonging to $\mathcal{F}$. If $\min\{X_1,\ldots,X_p\}$ and $\min\{Y_1,\ldots,Y_q\}$ have identical distributions, then $p = q$ and there exists a permutation $\{k_1,k_2,\ldots,k_p\}$ of $\{1,2,\ldots,p\}$ such that the pdf of Y_i if f_{k_i} $(i = 1,2,\ldots,p)$. Anderson and Ghurye (1977) proved a similar theorem for the maximum. As an application of the above theorem, consider the following examples.

Example 2. <u>$\mathcal{F}$ is the family of normal distribution</u>

$$\phi(x|\mu,\sigma) = \frac{1}{\sqrt{\pi}\sigma}\exp[-\tfrac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

Now,

$$\lim_{x\to-\infty}\frac{\phi(x|\mu_2,\sigma_2)}{\phi(x|\mu_1,\sigma_1)} = \begin{cases} 1, \text{ if } \mu_1 = \mu_2,\ \sigma_1 = \sigma_2 \\ 0, \text{ if } \sigma_2 < \sigma_1 \text{ or } \sigma_1 = \sigma_2 \text{ and } \mu_2 > \mu_1 \\ \infty, \text{ if } \sigma_2 > \sigma_1 \text{ or } \sigma_1 = \sigma_2 \text{ and } \mu_2 < \mu_1. \end{cases} \quad (7)$$

Conditions of the above theorem are met. Hence the distributions are identifiable.

Example 3. F is the family of negative exponential distributions

$$f_\lambda(x) = \lambda e^{-\lambda x}, \; x \geq 0, \; \lambda > 0$$

Here

$$\lim_{x \to 0} \frac{f_{\lambda_2}(x)}{f_{\lambda_1}(x)} = \begin{cases} 1, & \lambda_1 = \lambda_2 \\ \frac{\lambda_2}{\lambda_1} & \end{cases} \tag{8}$$

Here conditions of the above theorem are not met and hence the distributions are not identifiable.

Note, however, if the maximum is observed both normal and exponential distributions are identifiable.

There may, however, be situations when the conditions of the above theorem will not be met yet the underlying family of distributions may be identifiable. For example, Basu and Ghosh (1980) have shown that the gamma and the Weibull distribution, which contain the exponential distribution as special cases, are not identifiable. In fact they prove the following theorems.

Theorem 3. Let the pdf of X_i be given by

$$f_i(x) \equiv f(x;\alpha_i,\beta_i) = \frac{e^{-x/\beta_i} \, x^{\alpha_i - 1}}{\beta_i^{\alpha_i} \, \Gamma(\alpha_i)}, \; \alpha_i > 0, \; \beta_i > 0 \quad (i = 1,2,3,4)$$

where α_1 and α_2 are not both equal to one and α_3 and α_4 are not both equal to one. Let X_1 and X_2 be independent random variables. Similarly, let X_3 and X_4 be independent. If the distribution of

$\min(X_1,X_2)$ is identical with that of $\min(X_3,X_4)$, then either

$$(\alpha_1,\alpha_2) = (\alpha_3,\alpha_4) \text{ and } (\beta_1,\beta_2) = (\beta_3,\beta_4)$$

or,

$$(\alpha_1,\alpha_2) = (\alpha_4,\alpha_3) \text{ and } (\beta_1,\beta_2) = (\beta_4,\beta_3).$$

Now consider the case of the Weibull distributions.

Let $f_i(x) = \frac{p_i}{\theta_i} x^{p_i} \quad e^{-x^{p_i}/\theta_i}$, $x > 0$, $(\theta_i, p_i > 0)$.

Here $\bar{F}_i(x) = 1 - F_i(x) = e^{-x^{p_i}/\theta_i}$, $f_i(x)/\bar{F}_i(x) = \frac{p_i}{\theta_i} x^{p_i - 1}$, and as $x \to 0$,

$$\frac{f_i(x)}{f_j(x)} \longrightarrow \begin{cases} \theta_j/\theta_i, & p_i = p_j \\ 0, & p_i > p_j \\ \infty, & p_i < p_j \end{cases} \tag{9}$$

<u>Theorem 4</u>. Let $X_i \sim W(p_i,\theta_i)$, $(i = 1,2,3,4)$ be independent Weibull random variables. If the distribution of $\min(X_1,X_2)$ is the same as that of $\min(X_3,X_4)$, then either

$$(p_1,\theta_1) = (p_3,\theta_3),\ (p_2,\theta_2) = (p_4,\theta_4),$$

or,

$$(p_1,\theta_1) = (p_4,\theta_4) \text{ and } (p_2,\theta_2) = (p_3,\theta_3)$$

provided $p_1 \neq p_2$.

In Theorem 4 we exclude the case $p_1 = p_2$. For if $p_1 = p_2 = p$, say, the problem, after using the transformation $Y_i = X_i^p$ $(i = 1,2)$, reduces to that of the exponential distribution.

Note that the above suggests that Theorem 2, in its present form, provides a sufficient condition for identifiability. However, it is not a necessary one.

4. DEPENDENT RANDOM VARIABLES

We next consider the case when the random variables $(X_1, X_2, \ldots, X_p)$ are dependent. We restrict our discussion primarily to the case when $p = 2$. There are many physical situations where it is desirable to test the assumption so frequently made that the X_i's are independent. It is therefore natural to study the extent to which (U,I) or U determines the joint distribution of X's can be identified. To this end Basu and Ghosh (1978) pointed out the difficulties using the following construction. Let $F(x_1, x_2)$ be the joint distribution of (X_1, X_2), $\overline{F}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$, and $\overline{F}_i(x_1, x_2) = \partial \overline{F}(x_1, x_2)/\partial x_i$, $(i = 1,2)$. For simplicity assume that the density $f(x_1, x_2) > 0$ for all (x_1, x_2). Let

$$\overline{G}_i(x) = \exp\left\{ -\int_{-\infty}^{x} -\overline{F}_i(Z,Z)(\overline{F}(Z,Z))^{-1} dZ \right\} \tag{10}$$

and assume $\int_{-\infty}^{\infty} -\overline{F}_i(Z,Z)(\overline{F}(Z,Z))^{-1} dZ$ diverges for $i = 1,2$. Then $G_i(x) = 1 - \overline{G}_i(x)$ is a distribution finction and (U,I) has the same distribution whether (X_1, X_2) is distributed according to $F(x_1, x_2)$ or $G(x_1, x_2) = G_1(x_1) \cdot G_2(x_2)$. Thus our porblem could have a satisfactory solution only if F is known to be a well specified parametric family of distributions. Similar results for nonidentifiability, in the absence of specific parametric models, has also been considered by Miller (1977), Tsiatis (1975), and Rose (1973). Tsiatis (1978) further illustrated the magnitude of this problem with some actual data.

No general result on identifiability for dependent random variables is currently available. However Peterson, Jr. (1975) obtained the following interesting inequalities when $p = 2$. Let

$$\begin{aligned} Q_1(x_1) &= P(X_1 > x_1 \text{ and } X_1 < X_2), \\ Q_2(x_2) &= P(X_2 > x_2 \text{ and } X_2 < X_1), \\ Q_1(0) &= P(X_1 < X_2) = P_1, \\ Q_2(0) &= P(X_2 < X_1) = P_2. \end{aligned} \tag{11}$$

Then

$$\begin{aligned} \overline{F}(x_1,x_2) &= Q_1(x_1) + Q_2(x_2) - P(x_1 < X_1 < X_2 < x_2), \text{ if } x_1 < x_2 \\ &= Q_1(x_1) + Q_2(x_2) - P(x_2 < X_2 < X_1 < x_1), \text{ if } x_1 > x_2, \end{aligned} \tag{12}$$

$$Q_1[\max(x_1,x_2)] + Q_2[\max(x_1,x_2)] \le \overline{F}(x_1,x_2) \le Q_1(x_1) + Q_2(x_2),$$

and

$$\begin{aligned} Q_1(x_1) + Q_2(x_1) &\le \overline{F}_1(x_1) \le Q_1(x_1) + p_2 \\ Q_1(x_2) + Q_2(x_2) &\le \overline{F}_2(x_2) \le p_1 + Q_2(x_2) \end{aligned} \tag{13}$$

Bivariate normal distribution

In case of specific parametric models questions on identifiability has been settled on a number of distributions. We summarize some of these results. Consider first the case of bivariate normal distribution. Let $(X_1,X_2) \sim BVN(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho_{12})$. Nádas (1971) showed that if $0 < \rho_{12} < 1$ the distribution of the identified minimum of a normal pair determines the distribution of the pair. Nádas' proof is not complete, however. Basu and Ghosh (1978) completed Nádas' proof and extended it to the case of nonidentified

minimum. Recently Gilliland and Hannan (1980) gave an elegant solution of the problem for the general case. Their results are given below.

Let $(X_1,X_2) \sim BVN(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$. Assume, without any loss of generality, that $\sigma_1 > \sigma_2$ or $\sigma_1 = \sigma_2$ and $\mu_1 \geq \mu_2$. Identification of parameters $\mu_1,\mu_2,\sigma_1,\sigma_2$, and ρ within this restricted family is equivalent to identification of parameters up to the switch of (μ_1,σ_1) and (μ_2,σ_2) in the unrestricted family. Consider the following subfamilies of bivariate normal distributions defined by the following additional restrictions on the parameters.

$$\begin{aligned}
&N_0: \quad \sigma_2 > 0,\ |\rho| < 1 \\
&N_1: \quad \sigma_1 > \sigma_2 = 0 \\
&N_2: \quad \sigma_2 > 0,\ \rho = -1 \\
&N_3: \quad \sigma_1 > \sigma_2 > 0,\ \rho = 1 \\
&N_4: \quad \sigma_1 = \sigma_2 \geq 0,\ \rho = 1 \text{ or undefined.}
\end{aligned} \tag{14}$$

Let $N = \bigcup_0^4 N_i$. Let F_U denote the distribution of $U = \min(X_1,X_2)$ when (X_1,X_2) has distribution F.

<u>Theorem 5</u>. If $F,G \in N_0$ and $F_U = G_U$, then $F = G$.

<u>Theorem 6</u>. Suppose that $F,G \in N$ and $F_U = G_U$. Then $F,G \in N_i$ for some $i = 0,\ldots,4$. If $F,G \in N_i$, $i = 0,\ldots,3$ then $F = G$. If $F,G \in N_4$, then F and G have the same values μ_1,σ_1, and σ_2 values but arbitrary $\mu_2 \geq \mu_1$ values.

No result for the general case is available as yet. However Basu and Ghosh (1978) showed the identifiability of the trivariate normal distribution given the distribution of the identified minimum and, given that for each pair of random variable X_i, X_j with correlation cofficient ρ_{ij} and deviations σ_i and σ_j,

$$1 - \rho_{ij}\, \sigma_i/\sigma_j > 0 \ (i,j = 1,2,3;\ i \neq j).$$

Note most of the identification problems considered so far are also valid when the maximum V, instead of the minimum U, is observable. For if $\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ is p variate normal, so is $-\underline{X}$ and

$$\max(X_1,\ldots,X_p) = -\min(-X_1,\ldots,-X_p),$$

and thus identification problem for the maximum can be restated in terms of the identification problem for the minimum. Also, any bivariate distribution obtained through strict monotone transformation of normal variables will be identifiable. The bivariate lognormal distribution is thus identifiable.

Next we consider identifiability of a number of bivariate distributions useful in reliability theory and survival analysis. We, in particular, consider several bivariate exponential distributions. A survey of some of these distributions is presented in Basu and Block (1975). These include the bivariate exponential distributions of Marshall and Olkin (1967), Block and Basu (1974), and Gumbel (1960). Basu and Ghosh (1978, 1980) have considered identifiability of these distributions. There results are summarized below.

(a) Marshall and Olkin bivariate exponential. The tail probability of the Marshall-Olkin bivariate distribution is given by

$$\begin{aligned}\bar{F}(x_1,x_2) &= \exp[-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \max(x_1,x_2)], \\ &\qquad x_1,x_2,\lambda_1,\lambda_2 > 0, \quad \lambda_{12} \geq 0 . \\ &= 0, \text{ otherwise.}\end{aligned} \tag{15}$$

Here all parameters are identifiable if (U,I) is observed. However, if only U is observed the parameters are not identifiable.

(b) Block-Basu model. Here the joint density function is given by

$$\begin{aligned}f(x_1,x_2) &= \left\{\lambda\lambda_1(\lambda_2 + \lambda_{12})/(\lambda_1 + \lambda_2)\right\} \exp\left\{-\lambda_1 x_1 - (\lambda_2 + \lambda_{12})x_2\right\} \\ &\qquad x_1 < x_2 \\ &= \left\{\lambda\lambda_2(\lambda_1 + \lambda_{12})/(\lambda_1 + \lambda_2)\right\} \exp\left\{-(\lambda_1 + \lambda_{12})x_1 - \lambda_2 x_2\right\}, \\ &\qquad x_2 > x_1\end{aligned} \tag{16}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$. Here the parameters are not identifiable at all. Note that the model proposed by Freund (1961) is also not identifiable, since it is related to the Block-Basu model.

Because of the underlying physical assumptions neither Marshall-Olkin nor Block-Basu is considered a suitable one as a physical model when the maximum is observed.

(c) Gumbel model I. Gumbel (1960) proposed two bivariate exponential distributions. The first one is given by

$$F(x_1,x_2) = 1 - \exp(-\lambda_1 x_1) - \exp(-\lambda_2 x_2) + \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} x_1 x_2),$$
$$x_1, x_2, \lambda_1, \lambda_2 > 0\ , \quad \lambda_{12} \geq 0. \tag{17}$$

Here, the parameters are identifiable if (U,I) is observed. However, if the nonidentified minimum U is observed only λ_{12} and $\lambda_1 + \lambda_2$ are identifiable. If the nonidentified maximum $V = \max(x_1, x_2)$ is observable, then λ_{12} is identifiable and $(\lambda_1, \lambda_{12})$ is identifiable up to a permutation.

(d) Gumbel model II. Here the distribution function is given by

$$F(x_1,x_2) = (1 - \exp -\lambda_1 x_1)(1 - \exp - \lambda_2 x_2)(1 + \lambda_{12}\exp(-\lambda_1 x_1 - \lambda_2 x_2)) \tag{18}$$

Here if U is observable, λ_{12} is identifiable and (λ_1, λ_2) is identifiable up to a permutation.

(e) Bivariate Weibull distribution. Like the bivariate exponential distribution, the bivariate Weibull distribution can be defined in a number of ways. Consider the following bivariate Weibull survival function

$$\begin{aligned} \overline{F}(x_1,x_2) &= P(X_1 > x_1,\ X_2 > x_2) \\ &= \exp -\left\{\lambda_1 x_1^{p_1} + \lambda_2 x_2^{p_2} + \lambda_{12} \max(x_1^{p_1}, x_2^{p_2})\right\} \end{aligned} \tag{19}$$

Here again if U is observable, p_1 and p_2 are identifiable up to permutation. Also λ_1 and $\lambda_2 + \lambda_{12}$ or λ_2 and $\lambda_1 + \lambda_{12}$ are identifiable.

5. ESTIMATION OF PARAMETERS

Estimation of parameters based on (U,I), the identified minimum has been considered extensively. Basu and Ghosh (1978) have considered the problem of estimation of parameters based on U alone for the bivariate normal distribution. Similar results can be obtained using V.

We consider some other models. The pdf of V, assuming independence, is given by

$$f_V(t) = f_1(t)F_2(t) + f_2(t)F_1(t). \tag{20}$$

The parameters can therefore be estimated numerically using method of maximum likelihood.

As an illustration, let the density function of X_i be given by

$$f_i(x_i) = \lambda_i \exp(-\lambda_i x_i),\ x_i, \lambda_i > 0,\ (i = 1,2).$$

Then the density function of V is given by

$$f_V(t) = \lambda_1 \exp(-\lambda_1 t) + \lambda_2 \exp(-\lambda_2 t) - (\lambda_1 + \lambda_2) \exp -(\lambda_1 + \lambda_2)t.$$

Hence one can find the maximum likelihood estimates. A simpler method of estimation is by the method of moments. We can readily show

$$E(V) = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2} \tag{21}$$

and

$$E(V^2) = \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} - \frac{2}{(\lambda_1 + \lambda_2)^2}. \tag{22}$$

Replacing $E(V)$ and $E(V^2)$ by their simple estimates V_i/n and V_i^2/n one can solfe for $_1$ and $_2$ by iteration.

Similarly, suppose X_1, X_2 are dependent and follow the following distribution of Gumbel.

$$F(x_1,x_2) = 1 - e^{-\lambda_1 x_1} - e^{-\lambda_2 x_2} + e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} x_1 x_2}.$$

Here the density function of the maximum V is given by

$$f_V(t) = \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2 + 2\lambda_{12} t)e^{-(\lambda_1+\lambda_2)t - \lambda_{12} t^2} \quad (23)$$

Using this one can again estimate the parameters by the method of moments.

6. CONCLUDING REMARKS

In this section we point out a number of areas in which additional work is being carried out and we also point out a few open problems.

So far we have assumed, in the competing risks problem, that the observed random variable U is the minimum of p (unobserved) random variables $X_1, X_2, \ldots, X_p$. In analyzing mortality data the above is interpreted as follows. Consider a population in which K causes of death, $C_1, C_2, \ldots, C_p$, are operating. Each individual in this population is exposed to risk of dying of any one of these causes. One can recognize two kinds of distributions associated with death due to cause C_i:

(a) the survival distribution $\bar{F}_{ia}(t)$ due to cause C_i, conditionally that C_i is the cause of death, in the presence of other causes;

(b) the survival distribution $\bar{F}_i(t)$ due to C_i, if C_i is acting alone.

It is tacitly assumed that corresponding forces of mortality (failure rates) are the same. Gail (1975), Elandt-Johnson (1976), and others have explored the implication of this assumption.

A second assumption is that the potential survival times X_i's are independently distributed with continuous distribution function F_i $(i = 1,2,\ldots,p)$. Some results to this end are given in sections 2, 3, and 4. For another interesting direction of research see Miller (1977), Desu and Narula (1977), Langberg, Proschan, and Quinzi (1977) (1978), and the references therein. Desu and Narula consider the problem of estimating the distribution function $F_i(t) = P(X_i \le t)$

and provides sufficient condition on the distribution of $(X_1, \ldots, X_p)$ under which such an estimation is possible.

A third direction of research has been the interpretation of competing risks problems in terms of some stochastic process. Chiang (1968) has studied the problem of competing risks using time-nonhomogeneous Markov processes. Clifford (1977) and Berlin, Brodsky, and Clifford (1977) have considered the problem of identifiability for this situation.

The problems described in the previous chapters can be extended in several directions. The author is currently working on some of the problems stated below.

(a) As mentioned before, Theorem 2 of Section 3 does not provide a necessary and sufficient condition. It would be desirable to improve on this result.

(b) Most of the identifications results obtained so fare are for the case of two competing causes. These need to be generalized for the case of any number of variables. Some results to this end have been obtained by Basu and Ghosh (1980b). Necessary algorithms for estimating the parameters should also be obtained.

(c) The concept of competing risks is well known. Basu and Ghosh (1980) first coined the term <u>complementary risks</u> for the dual problem. In reliability theory the corresponding problems are for series and parallel systems It is natural to pose the following general problem corresponding to a k-out-of p system $(k \leq p)$. Recall a system is called k-out-of-p if the system operates so long as k or more components function. Let $X_{(r)}$ be the r^{th} ordered statistic among $X_1, X_2, \ldots, X_p$. Suppose only $X_{(r)}$ is observable, where $r = p - k + 1$.

Given the distribution function of $X_{(r)}$ could we uniquely determine the marginal distribution function F_i of X_i $(i = 1, 2, \ldots, p)$? Note, if $r = 1$, we obtain the case of competing risks. And if $r = p$, we obtain the case of complimentary risks. Below we get some partial solution for the case of three independent exponentially distributed random variables.

Let X_1, X_2, X_3 be three independent random variables and let $X_i \sim e(\lambda_i)$ $(i = 1, 2, 3)$. Suppose distribution of $X_{(2)}$ is known and it is known which of the three variables X_1, X_2, and X_3 is $X_{(2)}$. That is we assume $X_{(2)}$ to be <u>identified</u>. If the X_i's have a common distribution F with density function f we can again solve the problem readily. For

$$\bar{F}_{(2)}(t) = P(X_{(2)} > t) = 2\left(\frac{1}{6} - \frac{F^2(t)}{2} + \frac{F^3(t)}{3}\right) \tag{24}$$

In general, if X_i has distribution function F_i and density function f_i we have

$$f_{(2)}(t) = (F_1(t)f_2(t)\bar{F}_3(t) + (\bar{F}_1(t)f_2(t)F_3(t) \tag{25}$$

For the exponential case we have the following theorem.

<u>Theorem 7</u>. Let X_1, X_2, X_3 be independent random variables with $X_i \sim e(\lambda_i)$ $(i = 1, 2, 3)$. Similarly let X_1', X_2', X_3' be independent random variables with $X_i' \sim e(\lambda_i')$ $(i = 1, 2, 3)$. If the distribution of the <u>identified</u> r^{th} ordered statistics $U_{(r)}$ (among X_1, X_2, X_3) and $U_{(r)}'$ (among X_1', X_2', X_3') are identical then either $(\lambda_1, \lambda_2, \lambda_3) = (\lambda_1', \lambda_2', \lambda_3')$ or $(\lambda_1, \lambda_2, \lambda_3) = (\lambda_3', \lambda_2', \lambda_1')$.

Since the result is well known for $r = 1$ and $r = 3$ consider the case when $r = 2$.

Proof. Since the order statistics is "identified" let us assume $U_{(2)} = X_2$ and $U'_{(2)} = X'_2$. Now the densities of $U_{(2)}$ and $U'_{(2)}$ are identically equal. Hence

$$\lambda_2 e^{-\lambda_2 t}[e^{-\lambda_1 t} + e^{-\lambda_3 t} - 2e^{-(\lambda_1+\lambda_3)t}] \tag{26}$$

$$= \mu_2 e^{-\mu_2 t}[e^{-\mu_1 t} + e^{-\mu_3 t} - 2e^{-(\mu_1+\mu_3)t}] \quad \text{for all } t.$$

or,

$$\frac{\lambda_2}{\mu_2} e^{\{(\lambda_1+\lambda_2)-(\mu_1+\mu_2)\}t} = \frac{1 + e^{-(\lambda_3-\lambda_1)t} - 2e^{-\lambda_3 t}}{1 + e^{-(\mu_3-\mu_1)t} - 2e^{-\mu_3 t}}$$

$$\text{for all } t. \tag{27}$$

For simplicity assume $\lambda_1 < \lambda_3$ and $\mu_1 < \mu_3$. Then as $t \to \infty$, the right hand side of (27) tends to 1. Hence the left hand side limit of (27) is also 1. This implies $\mu_1 = \lambda_1$ and $\mu_2 = \lambda_2$. Substituting these values back in (26) we note $\mu_3 = \lambda_3$.

More results in this direction have been obtained by Basu and Ghosh (1980b).

REFERENCES

Anderson, D. R. and Burnham, K. P. (1976). Population ecology of the mallard: VI. The effect of exploitation on survival. United States Department of the Interior Fish and Wildlife Service, Resource Publication 128.

Anderson, T. W. and Ghurye, S. G. (1977). Identification of parameters by the distribution of a maximum variable. Journal of the Royal Statistical Society, Series B, 39, 337-342.

Basu, A. P. and Block, H. W. (1975). On characterizing univariate and multivariate exponential distributions with applications. Statistical Distributions in Scientific Work, 3, 399-421. G. P. Patil, S. Kotz, and J. K. Ord, eds. D. Reidel Publishing Company, Dordrecht-Holland.

Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model. Journal of Multivariate Analysis, 8, 413-429.

Basu, A. P. and Ghosh, J. K. (1980a). Identifiability of distributions under competing risks and complementary risks model. Communications in Statistics (to appear).

Basu, A. P. and Ghosh, J. K. (1980b). Unpublished manuscript.

Berlin, B., Brodsky, J., and Clifford, P. (1977). Testing disease dependence in survival experiments with serial sacrifice. (Unpublished)

Berman, S. M. (1963). Notes on extreme values, competing risks, and semi-Markov processes. Annuals of Mathematical Statistics, 34, 1104-1106.

Birnbaum, Z. W. (1979). On the mathematics of competing risks. DHEW Publication Number: (PHS) 79-1351, National Center for Health Statistics, Hyattsville, Maryland.

Block, H. W. and Basu, A. P. (1974). A continuous exponential extension. Journal of the American Statistical Association, 69, 1031-1037.

Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics. J. Wiley, New York.

Clifford, P. (1977). Nonidentifiability in stochastic models of illness and death. Proceedings of the National Academy of Science, 74, 1338-1340.

David, H. A. and Moeschberger, M. L. (1978). The Theory of Competing Risks. Charles Griffin and Company, London.

Desu, M. M. and Narula, S. C. (1977). Reliability estimation under competing causes of failure. The Theory and Applications of Reliability, 2, 471-481. C. P. Tsokos and I. N. Shimi, eds. Academic Press, New York.

Elandt-Johnson, R. (1976). Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. Scandinavian Actuarial Journal, 1, 37-51.

Freund, J. E. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association, 56, 971-977.

Gail, M. (1975). A review and critique of some models used in competing risks analysis. Biometrics, 31, 209-222.

Gilliland, D. C. and Hannan, J. (1979). Identification of the ordered bivariate normal distribution by minimum variate. To be published.

Langberg, N., Proschan, F., Quinzi, A. J. (1977). Transformations yielding reliability models based on independent random variables: A survey. Applications of Statistics, 323-337. P. R. Krishnaiah, ed. North Holland Publishing Company, Amsterdam.

Langberg, N., Proschan, F., Quinzi, A. J. (1977). Converting dependent models into independent ones, preserving essential features. nnuals of Probability, 6, 174-181.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. Journal of the American Statistical Association, 62, 30-44.

Miller, D. R. (1977). A note on independence of multivariate lifetimes in competing risks model. Annuals of Statistics, 5, 576-579.

Nadás, A. (1970). On estimating the distribution of a random vector when only the smallest co-ordinate is obserable. Technometrics, 12, 923-924.

Nadás, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. Technometrics, 13, 201-204.

Peterson, A. V., Jr. (1975). Bounds for a joint distribution function with fixed sub-distribution functions; application to competing risks. Technical Report Number 7, Stanford University, Division of Biostatistics.

Puri, P. S. (1979). On certain problems involving non-identifiability of distributions arising in Stochastic modeling. Proceedings of a Symposium on Optimizing Methods in Statistics, 403-417. J. S. Rustagi ed. Academic Press, New York.

Rose, D. M. (1973). An investigation of dependent competing risks. Unpublished doctoral thesis, Univeristy of Washington.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Science, 72, 20-22.

Tsiatis, A. (1978). An example on nonidentifiability in competing risks. Scandinavian Actuarial Journal, 235-239.

DAT
ILM